
Customer Behavior Analysis on a Tmall E-commerce Shop

Renhao Jin, Song Han, Tao Liu, Songnan Xi

School of Information, Beijing Wuzi University, Beijing, China

Email address:

Renhao.jin@outlook.com (Renhao Jin)

To cite this article:

Renhao Jin, Song Han, Tao Liu and Songnan Xi. Customer Behavior Analysis on a Tmall E-commerce Shop. *International Journal on Data Science and Technology*. Vol. 2, No. 6, 2016, pp. 57-61. doi: 10.11648/j.ijdst.20160206.11

Received: October 5, 2016; **Accepted:** November 11, 2016; **Published:** November 25, 2016

Abstract: In recent years, China online marketing is very hot, and a lot of online shops run in the Tmall.com. This paper does an analysis of customer shopping behaviors in a certain e-commerce shop in Tmall. The shop is named X in this paper for privacy. Based on the descriptive analysis, it finds the profit customers and profit products in the shop. Then K-mean segmentation method is used to class the customers into 4 groups and the profiles of the customers in each group are described. The results on this paper can help the X shop to offer good services for the profit customers and do the precision marketing for all customers.

Keywords: E-commerce, CRM, K-mean Segmentation, Cluster Profile

1. Introduction

Tmall.com, formerly Taobao Mall, is a Chinese-language website for business-to-consumer (B2C) online retail, spun off from Taobao, operated in China by Alibaba Group. It is a platform for local Chinese and international businesses to sell brand name goods to consumers in mainland China, Hong Kong, Macau and Taiwan. Tmall.com currently features more than 70,000 international and Chinese brands from more than 50,000 merchants and serves more than 180 million buyers. Tmall.com ranked number one among all Chinese B2C retail websites for 2010 in terms of transaction volume, with a gross merchandise volume of RMB30 billion – about three times the amount facilitated by 360buy, its closest competitor. The site accounts for a 47.6% share of the B2C online retail market in China.

The data in this paper is from a shop in running on Tmall.com. This shop wants to analyze its customer shopping data to better service their customers and improve the shop profit. Customer behavior analysis is the important part in Customer relationship management. The customer behavior analysis tries to analyze data about customers' history with a company, to improve business relationships with customers, specifically focusing on customer retention, and ultimately to drive sales growth.

The X e-commerce shop now sells products in 5 categories of Homeware, Car supplies, Electronics, Kitchenware, and Beauty, and their customers spread in all provinces in China.

The data of this study includes everyday trade history for each customer from July 16th, 2015 to Oct. 28th, 2015. Seven variables are in the data, i.e., Customer ID, Product Category, Product Quantity, Product Price, Sale Amount, Sale Date and Customer Area. This paper is to analyze these customer behavior data. According to the results of the analysis, it is expected to give some advices on the business direction, sale strategies and future development. Hopefully, the analysis results can help to do precision marketing for customers, and lead more income for the X e-commerce shop.

2. K-means Segmentation

K-means segmentation is a widely used fast cluster method, and also named K-means method. It was originally proposed as a heuristic algorithm for finding clusters rather than as a formal statistical model. A lot of statistical or data mining textbooks include the detailed introduction on K-means method, thus the authors do not explain this method here.

In the data, there are 4862 customers and 6442 shopping records, with Seven variables in the data, i.e., Customer ID, Product Category, Product Quantity, Product Price, Sale Amount, Sale Date and Customer Area. The data are then combined for segmentation, with each customer has a summarized observation. The summarized observations are used to K-Means cluster, and variables in the summarized observations include total consumption amounts, total shopping times, total respective shopping times in Homeware, Car supplies, Electronics, Kitchenware, and Beauty. These

total variables for each customer are combined from the original data. The data are transformed by z-score standardized before clustering.

By clustering, the 4862 customers are grouped into several clusters, and the analysis on the profile of these clusters is then performed.

3. Descriptive Study

Figure 1 shows the spatial distribution of sales amount in each province. The name and sale amount of each community are marked in the map. The color of each community is proportional to total sales amount. Through the picture, it clearly displays the sales situations in all around China. Jiangsu has the highest sale amount, followed by Shandong, Anhui, Beijing and other places. These regions with high sale amounts are all in the coastline areas, and these areas are well developing in China. The residents in these areas have more money for shopping. Anhui province is not a rich area in China but has high sale amount, as the X shop is located in Anhui. For the other internal areas, their logistics systems are not well developed, and people prefer shopping underline to online. The incomes of residents in internal areas are not high. All the reasons make the sale amount in these areas are a little low. The X shop should put more advertisements in coastline areas and storage products suitable for the residents in coastline areas.

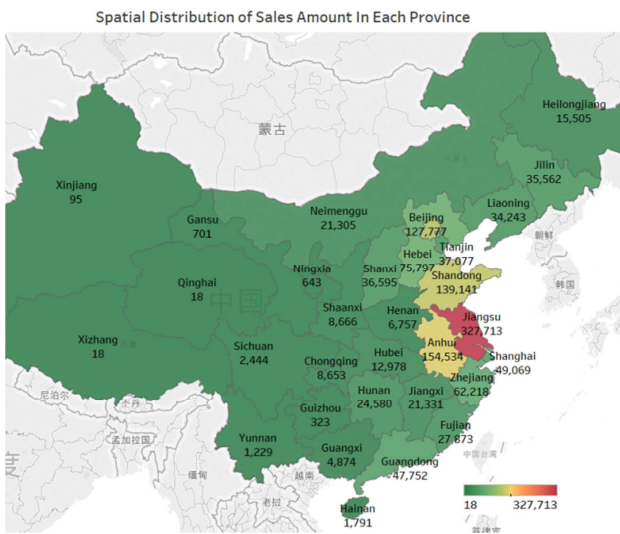


Figure 1. The spatial distribution of sales amount in each province. The name and sale amount of each community are marked in the map. The color of each community is proportional to total sales amount.

The scatterplot of product price and its quantity for each product is shown in Figure 2. Each color and each shape indicates each product type as shown in the figure index. The products with high price and high sale quantity are high profit products, and they locate in the up-right areas in the Figure 2. The X shop has the two 100 rules for high profit products, i.e., products with price higher than 100 and quantity more than 100. It is easy to list all the high profit products from the figure 2, but the lists are not shown here as all Chinese things. From

figure 2, it also can be found that many profit products are in Homeware category, and fewer in Car Supplies and Kitchenware category. From these results, the X shop can change their storage structure to better fit the needs of their customers.

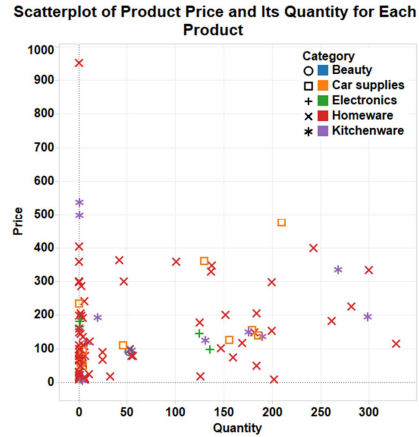


Figure 2. The scatterplot of product price and its quantity for each product. Each color and each shape indicates each product type as shown in the figure index.

Figure 3 displays the scatterplot of total shopping amount and total shopping times for each customer. Each circle indicates a customer, and the color of each circle is proportional to its total shopping amount. The customers with high total shopping amount and high shopping times are high profit and good customers, and they locate in the up-right areas in the Figure 3. The X shop has a simple rule for high profit customers, i.e., customers with total shopping amount higher than 1000 and shopping quantity more than 5 times. It is easy to list all the high profit products from the figure 3, but the lists are not shown here for privacy. From these results, the X shop should put much resources on better service to high profit customers.



Figure 3. The scatterplot of total shopping amount and total shopping times for each customer. Each circle indicates a customer, and the color of each circle is proportional to its total shopping amount.

4. Clustering Result

The original data are manipulated to 4862 observations with seven variables, i.e., total consumption amounts, total shopping times, total respective shopping times in Homeware, Car supplies, Electronics, Kitchenware, and Beauty. Each observation indicates the summary of a customer. By K-Means clustering, the 4862 customers are grouped into 4 clusters, and the segmentation results are shown in Figure 4.

In every plot of figure 4, the red bar chart is the distribution for all customers, while the blue chart is the distribution for the customers in the current cluster.

As shown in the Figure 4, Cluster 3 have largest customers with 3136, followed by cluster 1 with 924, Cluster 4 with 692, and Cluster 2 with 110.

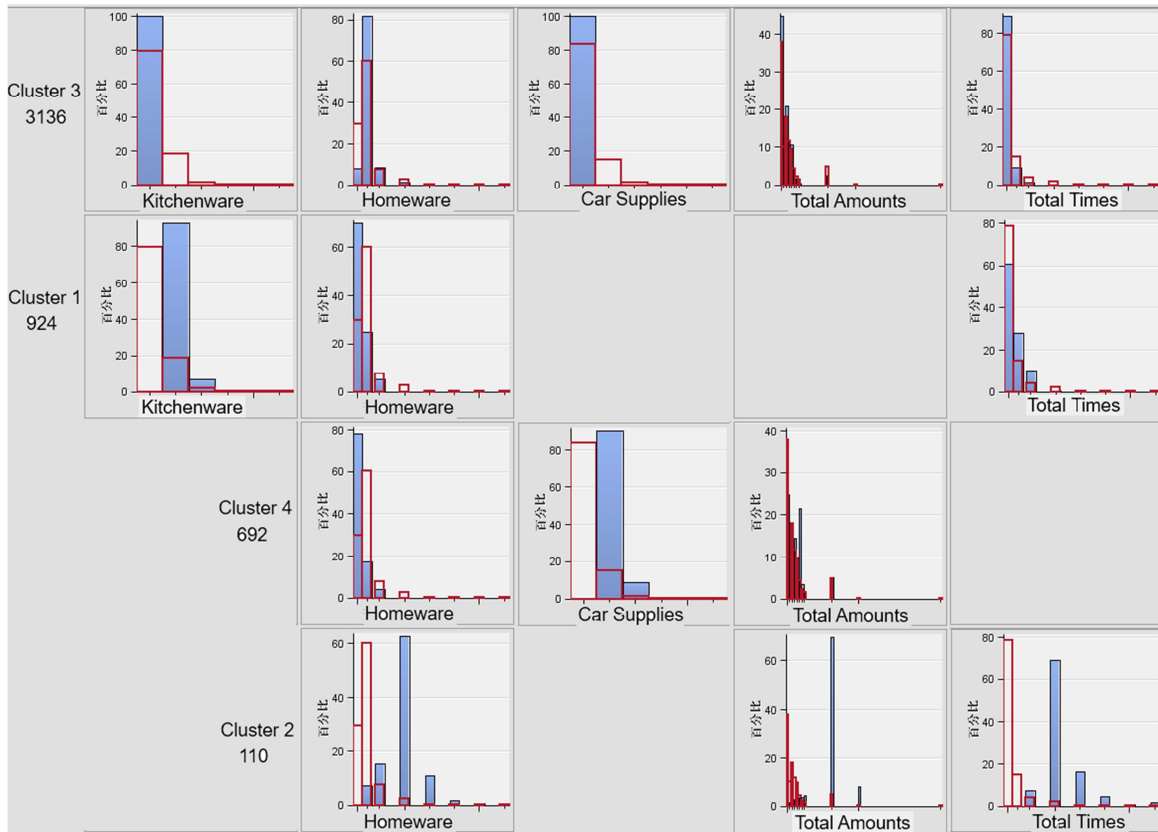


Figure 4. The profile for each cluster. In every plot, the red bar chart is the distribution for all customers, while the blue chart is the distribution for the customers in the current cluster.

From the Figure 4, it can be seen that customers in Cluster 3 have small shopping times in Homeware, Car supplies, and Kitchenware, and also have small total shopping amount and total shopping times. It means the customers in this cluster have no preference on shopping, and they are random customers. They are not positive on shopping, but at the same time they are also the potential customers, and the X shop should analyze these customers detailed to improve their shopping amounts and times.

For customers in Cluster 1 and Cluster 4, they are in the median place of total shopping amounts and shopping times. They are median positive customers but they have their own preferences from the Figure 4. For example, customers in Cluster 1 have more shopping times in Kitchenware category, while customers in Cluster 4 have more shopping times in Car Supplies category. So the X shop should advertise more Kitchenware products for customers in Cluster 1, and more Car Supplies products for customers in Cluster 4.

For customers in Cluster 2, they have higher total shopping

amount and higher total shopping times. They are high profit customers. From the Figure 4, it can be found that these customers have preference on shopping, i.e., their shopping times in homeware are relative high. From these results, the X shop should put more resource on service for the customers in this Cluster, as high profit customer often can bring more profits. At the same time, the X shop should advertise more products on Homewares for these customers.

5. Conclusions

This paper does an analysis on customer behavior in an e-commerce shop in Tmall.com. The X e-commerce shop now sells products in 5 categories of Homeware, Car supplies, Electronics, Kitchenware, and Beauty, and their customers spreads in all provinces in China. It is found that Jiangsu has the highest sale amount, followed by Shandong, Anhui, Beijing and other places. The X shop should put more advertisements in coastline areas and storage products suitable

for the residents in coastline areas. The products in Kitchenware category are more welcomed by customers, and the detailed high profit products are not listed but the X shop can change their storage based on the list. The high profit customers are shown in Figure 3, and the X shop should put much resource on better service to high profit customers.

The 4862 customers are segmented into 4 clusters based on their shopping history. The profile of each cluster is analysed. The customers in Cluster 3 are not positive customers and no shopping preferences are found. The customers in Cluster 1 and Cluster 4 are median positive customers, and customers in Cluster 1 have more shopping times in Kitchenware category, while customers in Cluster 4 have more shopping times in Car Supplies category. For customers in Cluster 2, they have higher total shopping amount and higher total shopping times. They are high profit customers. The X shop should make different shopping service for these 4 groups of customers for precision marketing.

This paper does a simple description and analysis on the data from X platform, and the results in this paper can be a reference to X platform. Hopefully, the results can be used to modify the strategies of the X platform for future development.

Acknowledgements

This paper is funded by the project of National Natural Science Fund, Logistics distribution of artificial order picking random process model analysis and research (Project number: 71371033); and funded by intelligent logistics system Beijing Key Laboratory (No. BZ0211) and Beijing Intelligent Logistics System Collaborative Innovation Center; and funded by scientific-research bases---Science & Technology Innovation Platform---Modern logistics information and control technology research (Project number: PXM2015_014214_000001); University Cultivation Fund Project of 2014-Research on Congestion Model and algorithm of picking system in distribution center (0541502703).

References

- [1] An W, Liu Q H. Research on Apparel E-Commerce Customer Behavior Based on Life Style [J]. *Applied Mechanics & Materials*, 2013, 411-414: 2173-2176.
- [2] Bhate D V, Pasha M Y. Analyzing Target Customer Behavior Using Data Mining Techniques for E-Commerce Data [J]. *Ijircst Org*, 1970.
- [3] Bennett Horter. MANAGEMENT SUMMARY: A Phase I Archaeological Survey of the Proposed McGarry Hollow Flood Analysts in Weston, West Virginia [J]. *New England Journal of Medicine*, 1943, 300 (11): 619-20.
- [4] Crider, Franklin J. (Franklin Jacob). Establishing a commercial vineyard in Arizona [J]. *Establishing A Commercial Vineyard in Arizona*, 1923.
- [5] Chen F, Wang B, Wei Z. The rise of the internet city in China: Production and consumption of internet information [J]. *Urban Studies*, 2014, 52(13): págs. 2313-2329.
- [6] Denecker M, Marek V, Truszczyński M. Approximating operators, stable operators, well-founded fixpoints and applications in non-monotonic reasoning [J]. *Springer International*, 2000, 597: 127-144.
- [7] Calof J L. The Relationship Between Firm Size and Export Behavior Revisited [J]. *Journal of International Business Studies*, 1994, 25(2): 367-387.
- [8] Conrick R, Curtis N L, Staten P W, et al. The relationships between temperature gradient and wind during cold frontal passages in the eastern United States: a numerical modeling study [J]. *Atmospheric Science Letters*, 2016.
- [9] Dong L, Ye T. An Empirical Investigation of Chinese Online Consumer Ethics [J]. *Open Journal of Social Sciences*, 2015, 03(10): 161-169.
- [10] Dat, Le Ba, Viet, Nguyen Hoang, Trang, Nguyen Thi Huyen, et al. E-COMMERCE APPLY CUSTOMER'S BEHAVIOR ANALYSIS [J]. 2012.
- [11] Elahi S. A Survey on the Relationship Between E-Commerce and Customer Behavior [J]. *Business Strategies*, 2009.
- [12] Hirsch J, Jr A E. The separation of complex lipide mixtures by the use of silicic acid chromatography. [J]. *Journal of Biological Chemistry*, 1958, 233(233): 311-320.
- [13] Hiller D M, Martin D S. Radiochemical Studies on the Photofission of Thorium [J]. *Physical Review*, 1953, 90(4): 581-585.
- [14] He L, Xie X, Jin H, et al. Harnessing Dynamic Interests of Crowd in Chinese Online Shopping Festivals [C]// *IEEE, Computer Software and Applications Conference*. IEEE Computer Society, 2015: 806-815.
- [15] Jiang Y, Yu S. Mining E-Commerce Data to Analyze the Target Customer Behavior [C]// *International Workshop on Knowledge Discovery and Data Mining*. 2008: 406-409.
- [16] Labo, Jackson. Rapid endocytosis coupled to exocytosis in adrenal chromaffin cells involves Ca²⁺, GTP, and dynamin but not clathrin [J].
- [17] Lee H J. Effects of Product Recommendations on Customer Behavior in e-Commerce: An Empirical Analysis of Online Bookstore Clickstream Data [J]. 2008.
- [18] Liu Q, Zeng X, Liu C, et al. Mining Indecisiveness in Customer Behaviors [C]// *IEEE International Conference on Data Mining*. IEEE, 2015: 281-290.
- [19] Makori W. Real-time information processing and supply chain optimization among supermarkets in Nairobi, Kenya [J]. *Degree in Master of Business Administration*, 2013.
- [20] Nikhita M, Ashwini J, Amritha P, et al. MINING THE E-COMMERCE DATA TO ANALYZE THE TARGET CUSTOMER BEHAVIOR [J]. *Journal of Policy & Organisational Management*, 2014.
- [21] Nikhita M, Ashwini J, Amritha P, et al. Mining the E-commerce Data to Analyze the Target Customer Behavior [C]// *International Workshop on Knowledge Discovery and Data Mining*. 2008.
- [22] Pfeifer R, Stenzel G, Wolfram K. Behavior analysis based optimization of navigation in E-commerce user interfaces: US, US20040095383 [P]. 2004.

- [23] Rogers P G, Zuber M T. Tectonic evolution of Bell Regio, Venus: Regional stress, lithospheric flexure, and edifice stresses [J]. *Journal of Geophysical Research Atmospheres*, 1996, 27(E7): 54–57.
- [24] Rao Y V, Budde S R. E-commerce and customer behavior in India: Factor analysis [J]. 2015, 5(2).
- [25] Russell A R. *Unsaturated Soils: Research & Applications* [J]. Crc Press, 2014.
- [26] Smith R K, Reeder M J. On the Movement and Low-Level Structure of Cold Fronts [J]. *Monthly Weather Review*, 1988, 116(10): 1927-1944.
- [27] Sokolowski M, Koch T, Pfnür H. Ordered structures and phase diagram of atomic hydrogen chemisorbed on ruthenium (001)[J]. *Surface Science Letters*, 1991, 243(1-3): 261-272.
- [28] Schied A, Strehle E, Zhang T. High-frequency limit of Nash equilibria in a market impact game with transient price impact [J]. *Papers*, 2015.
- [29] Wohlsen M. Chinese Giant Alibaba Is Ready to Become the Next Google | WIRED [J]. 2014.
- [30] Xiong Y. Logistics Distribution Modes of B2C E-Commerce: a Case Study of Tmall and 360Buy [J]. *Business Economy*, 2013.
- [31] Yadav M P, Feeroz M, Yadav V K. Mining the customer behavior using web usage mining in e-commerce [C]// *International Conference on Computing Communication & NETWORKING Technologies*. 2012: 1-5.
- [32] Yi C. Enticing and Engaging Consumers via Online Product Presentations: The Effects of Restricted Interaction Design [J]. *Journal of Management Information Systems*, 2015, 31(4): 213-242.
- [33] Yang J, Luo J, Shen J, et al. Online Shopping Preference Analysis of Campus Network Users Based on MapReduce [C]// *International Conference on Cloud Computing and Big Data*. IEEE, 2014: 138-143.
- [34] Yingst R A, Iii J W H. Volumes of lunar lava ponds in South Pole-Aitken and Orientale Basins: Implications for eruption conditions, transport mechanisms, and magma source regions [J]. *Journal of Geophysical Research Atmospheres*, 1997, 102(E5): 10909-10931.